2. Description of Data

A total of three different data sources are required, data from venues, from New York's and Toronto's neighborhoods.

2.1. Venues of FourSquare

The first data source is FourSquare, this is also the main data source. The FourSquare database is used to obtain the advantages of a city. FourSquare is a location-based recommendation service in the form of application software for event locations such as restaurants. It also offers a free API for developers to access its databases. It contains global location data for 190 countries and territories. The venues are categorized, and the geographic coordinates can be accessed.

Above all, categories of event locations are required for the analysis. FourSquare has a total of ten categories, which are divided into a total of 470 sub-categories. The breakdown of the subcategories is as follows:

	Categorie	Amount of Subcategories
0	Arts & Entertainment	38
1	College & University	23
2	Event	12
3	Food	92
4	Nightlife Spot	7
5	Outdoors & Recreation	66
6	Professional & Other Places	44
7	Residence	5
8	Shop & Service	147
9	Travel & Transport	36

It can be seen in the graphic that the main categories are as follows:

> 'Arts & Entertainment', 'College & University', 'Event', 'Food', 'Nightlife Spot', 'Outdoors & Recreation', 'Professional & Other Places', 'Residence', 'Shop & Service', 'Travel & Transport'

2.2. Neighborhoods and coordinates of New York

The second data source is a JSON file that contains the names and coordinates of New York's neighborhoods. The names of geographical coordinates can be extracted and saved in a data frame. This is to provide an overview of Mr. Miller's current environment. There is a total of five boroughs with a total of 306 neighborhoods. The boroughs are the following:

> 'Bronx', 'Manhattan', 'Brooklyn', 'Queens', 'Staten Island'

A small overview of this looks like this:

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

In addition, the geographical coordinates of the centre of New York are required, which reflect the starting point and the current place of residence of Mr. Miller. These coordinates are obtained using the Python geopy library, which is a Python client for several popular geocoding web services.

Mr. Miller's coordinates are as follows: latitude = 40.7127281, longitude = -74.0060152

2.3. Neighborhoods and coordinates of Toronto

The third and final data source is Wikipedia. The encyclopedia provides an overview of the Toronto neighborhoods. This is to be the new home of Mr. Miller and his family. The URL is the following:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The table looks like this:

	Postal Code	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront

After cleaning up the data, the geographical coordinates of the neighborhoods can be added again. These can be determined and added using the zip code and the Python library "geocoder". Accordingly, the overview looks like this:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.8114	-79.1966
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.7857	-79.1587
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.7658	-79.1747
3	M1G	Scarborough	Woburn	43.7681	-79.2176
4	M1H	Scarborough	Cedarbrae	43.7694	-79.2389

Toronto has a total of ten boroughs, with a total of 103 neighborhoods. The ten boroughs are as follows:

> 'Scarborough', 'North York', 'East York', 'East Toronto', 'Central Toronto', 'Downtown Toronto', 'York' 'West Toronto', 'Mississauga', 'Etobicoke'

Finally, we need the geographic coordinates of the centre of Toronto, which is Mr. Miller's new place of residence, in which his new job is. These coordinates are obtained again with the Python library "geopy".

The coordinates of his new job look like this: latitude = 43.6534817, longitude = -79.3839347